Check Your Facts and Try Again: Improving Large Language Models with External Knowledge and Automated Feedback

Year: 2023 | *Citations: 475* | *Authors: Baolin Peng, Michel Galley, Pengcheng He*

Abstract

Large language models (LLMs), such as ChatGPT, are able to generate human-like, fluent responses for many downstream tasks, e.g., task-oriented dialog and question answering. However, applying LLMs to real-world, mission-critical applications remains challenging mainly due to their tendency to generate hallucinations and their inability to use external knowledge. This paper proposes a LLM-Augmenter system, which augments a black-box LLM with a set of plug-and-play modules. Our system makes the LLM generate responses grounded in external knowledge, e.g., stored in task-specific databases. It also iteratively revises LLM prompts to improve model responses using feedback generated by utility functions, e.g., the factuality score of a LLM-generated response. The effectiveness of LLM-Augmenter is empirically validated on two types of scenarios, task-oriented dialog and open-domain question answering. LLM-Augmenter significantly reduces ChatGPT's hallucinations without sacrificing the fluency and informativeness of its responses. We make the source code and models publicly available.